DOCUMENT RESUME

ED 228 287 TM 830 177

AUTHOR Williams, Richard C.; Bank, Adrianne
TITLE Linking Testing and Evaluation Activities with Instruction: Can School Districts Make It Happen?
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO CSE-R-184
PUB DATE 82
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Decision Making; Elementary Secondary Education; Evaluation; *Evaluation Utilization; Field Studies; *Instructional Development; Longitudinal Studies; *School Districts; *Testing; *Test Use

ABSTRACT

The attempt by a small number of school districts to forge a linkage among testing and evaluation and instructional decision making is examined. The sample districts (approximately six) exhibit characteristics that represent the diversity of American school districts in size, locale, and student demographics. Three researchers spent approximately 1 week in each district visiting schools and district offices, interviewing district participants, and examining relevant documents and records. District respondents were asked about three general areas: (1) Why is this district trying to link testing and evaluation with instructional decision-making? (2) How does this district do this? and (3) What effects have the linking activities had? The following three specific questions are discussed as they relate to the first two areas of interest: (1) What are the incentives and disincentives that operate in school districts attempting to forge an evaluation-testing-instruction linkage? (2) What are examples of the approaches districts are taking to forge these linkages? and (3) What are the potential contributions this research has for school improvement? (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Gray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED228287

LINKING TESTING AND EVALUATION ACTIVITIES WITH INSTRUCTION: CAN SCHOOL DISTRICTS MAKE IT HAPPEN?*

Richard C. Williams
Adrianne Bank

CSE Report No. 184
1982

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California, Los Angeles

*This paper was prepared for a symposium on NIE funded research on Testing and Evaluation in School Districts, American Educational Research Association, Los Angeles, 1981.

TM 820 177

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

In July 1979, we began a three-year inquiry to discover ways in which school districts might effectively link their district testing and evaluation activities with instructional decision making.

This inquiry was stimulated by our belief, based on previous research and experience in school districts, that testing and evaluation activities in most districts had only limited influence on internal school district instructional decision making. Instead, the focus of testing and evaluation in many districts seemed to be toward satisfying external demands, e.g., federal program evaluation requirements, court-ordered desegregation mandates (Zucker, 1981; David, 1978). But many school districts had moved to develop their testing and evaluation capacities (Lyon, et al., 1978) and it seemed logical to us that the data and reports generated by a district evaluation unit might also serve as a district curriculum and instructional management information system.

The main purpose of our work is not to determine the extent to which a nationwide sample of school districts are using testing and evaluation for internal instructional decision making. Instead, we are examining how a small number of districts are attempting to forge a linkage among testing and evaluation and instructional decision making.

At the present time, we have completed extensive case studies in five or six districts that we selected because they had a reputation for having tried to forge this linkage. Our sample districts, while not comprising a national sample, do exhibit characteristics that represent the diversity of American school districts. They reflect differences in:

size (large/small), student demographics (affluent/below-average income, racially homogeneous/racially heterogenious), and locale (urban/suburban). Three researchers have each spent approximately one week in each district visiting schools and district offices, interviewing district participants, examining relevant documents and records. We have asked respondants about three general areas: Why is this district trying to link testing and evaluation with instructional decision-making? How does this district do this? What effects have the linking activities had?

In the brief space available to us, we would like to discuss three specific questions related only to the first two areas of interest.

1. What are the incentives and disincentives that operate in school districts attempting to forge an evaluation-testing-instruction linkage?
2. What are examples of the approaches districts are taking to forge these linkages?
3. What are the potential contributions this research has for school improvement?

But before doing so, we'd like to define briefly what we mean by linkage although you will get its fuller flavor, by example, later in the paper. Linkage, to us, means the coordination--either through formal or informal means--of all the operations and services within a school district essential or supportive of the use of testing and evaluation for instructional purposes. Linkage is a function of management. It is an arrangement which brings together in some productive manner data collection, analysis, and reporting with core instructional activities.

Such testing-evaluation-instruction linkages are not commonplace in school districts although testing and evaluation activities have increased substantially since 1965. This may mean that most school districts have, over the past 15 years, felt little need to make such a linkage. We were interested to learn what factors seemed to be encouraging our sample districts to move in this direction.

Question 1. What are the incentives and disincentives that operate in school districts attempting to forge an evaluation-testing-instruction linkage?

In the districts we studied, the single shared reason given for initiating coordination arrangements between tests and evaluations was to influence pupil achievement as measured on test scores. In many of the districts there had been expressed dissatisfaction, coming from a number of sources, with the academic performance of students. The move towards use of tests and evaluation data was primarily remedial. In one of the districts, however, there had been overall satisfaction with student learning; moreover, there was a sense, on the part of the district superintendent, that individualized instruction might increase the learning of average and above average students.

District officials indicated in their interviews with us that their overall intention was to use test scores as a description of student achievement. They wanted these scores arranged and understood in such a way so as to redirect instruction. However, the immediate incentives for starting and continuing such a process seemed to vary from district to district. For example, some central offices were moved in this direction by explicit mandate from courts, or from state legislatures or from school boards. In other districts, superintendents or other officials seemingly influenced

by research and current educational thinking, decided to use available federal and state money to build instructionally relevant tests.

We might categorize the types of incentives we found as either "sticks" or "carrots" and their sources as either external or internal to district management. Our matrix would look something like this.

---

insert Figure 1 about here

This list of incentives, to some extent, begs the question. The carrots and the sticks are common to other districts. Why haven't they moved to link testing and evaluation with instruction? Given our small sample, and our field-based research design, we cannot provide a general answer to that question. What we can say is that certain characteristics seem to be present in our five districts, especially those that are most advanced in their linkage development. These elements indicate that our districts had the management capacity to respond to the incentives. The elements we refer to are: idea champions, stable core staff, realistic problem analysis, and tolerance for ambiguity. The following is a brief description of each element:

- Idea champions--by this we mean individuals in key administrative and policy positions who firmly believe in the value of test and evaluation data and consistently champion its development and connection to instruction. In our districts, these individuals were found in a variety of positions. There was no consistent pattern to their school district assignments, e.g., some are in curriculum, some in evaluation, some are line administrators; what they do share with

	Sticks	Carrots
External	°requirements by federal or state agencies to: °evaluate problems °develop courses of study °raise test scores °community dissatisfaction with public education expressed by: °press and media °loss of students °Board action	°availability of federal and state money for °text development °evaluation of programs °staff development °relationships with universities °existence of techniques or procedures to link tests with instruction
Internal	°decision of district administrators to link testing, evaluations, and instruction	°desire of district to acquire additional funds °presumed likelihood of success in linking testing and evaluation with education

Figure 1

one another are characteristics such as conviction, persuasiveness, and some degree of power.

- Stable core group--in our districts, these "idea champions" and their followers have been around for a while. In our most advanced linkage systems it has taken from 8-10 years for the linkage programs to develop and mature. This could not have happened if the core group had continually changed.
- Comprehensive rather than ad hoc problem analysis--the core group has been aware of next steps beyond the immediate task of the moment. It is one thing to develop, for example, a CRT program in reading--it is quite another thing to actually get teachers to use it. Bridging the gap between development and use implies an understanding of the school site and district as a bureaucratic social system and an appreciation of the various strategies and tools that might most effectively bridge the gap.
- Tolerance for ambiguity--none of the linkage arrangements developed, over time, in anything resembling the rational, linear way that is often described in standard planning texts. Instead, the programs have developed unevenly, component by component on a broken front. Many times, the components of the linking system have been developed independently of one another, with different purposes and each with its own set of advocates. Developing linkage arrangements to merge together these disparate pieces into a new configuration takes time and it can be very frustrating.

The above list is not exhaustive and it may be that these characteristics and activities are found in many districts that have not thought about or who have tried and abandoned an effort to develop a linkage system; we don't know. All we can say is that these are present in our sample districts and we believe they contribute to the progress these districts have made.

What about disincentives to develop a management arrangement that links services and supports to the connection between testing, evaluation, and instruction? As we indicated above, typically districts are not moving in this direction. There are likely several reasons for this. An important one, we believe, is that these districts are not pressured or pulled to think about the impact of students' test scores through change in instructional activities. Thus, they continue in a traditional arrangement of semi-autonomous operational units.

For example, districts may feel that their declining test scores are caused by large and rapid changes in the ethnic or racial class make-up of their pupil population or that their declining performance merely reflects the flagging public support for the schools. They reason that, until these conditions change, it is unlikely that encouraging curriculum and instructional changes based on test scores and evaluation findings will make much of a difference. They conclude that other political, social, or financial strategies might be more appropriate.

Another disincentive is that a closely linked testing, evaluation, and instructional system, with its emphasis on supervision, communication, and coordination, flies in the face of the traditional school district operating mode which can be characterized as loosely coupled,

(Weick, 1976) with teachers working quite independently behind closed classroom doors (Lortie, 1975). Teachers do not readily embrace approaches that fundamentally alter their accustomed professional behavior patterns.

Still another disincentive may be that a tight and interactive relationship between test scores and classroom practice is yet an unproven solution to the problem of student learning. While various components, e.g., development and use of CRTs, formative and summative evaluation methods, are becoming increasingly sophisticated and technically sound, much remains to be done before teachers and administrators are convinced that these techniques can be used as effective tools in their own classrooms for improving student achievement. Given the other demands on their time and energy, teachers will not readily commit themselves to unknown and unproven technologies.

Question 2. What are examples of the approaches districts are taking to forge these linkages?

Our sample districts are using a variety of approaches in linking testing and evaluation with instructional improvement. In this paper we will very briefly describe two approaches: a decentralized, school-oriented system using norm-referenced standardized test scores; and a district-directed centralized system using district-directed criterion-referenced tests. The decentralized NRT system uses the individual school as the locus of change. Within loosely prescribed district parameters, each school has considerable discretion in developing and implementing an instructional program that the school staff feels is appropriate for its particular student body. The norm-referenced student test results are folded into an individualized evaluation report that is prepared

for each school by the district office. The school staff, with the assistance of the central evaluation unit staff and, often, with district instructional and curricular specialists, develop yearly plans in which they identify their own instructional and other programmatic priorities. Presumably, the evaluation unit's reports, including the scores, form part of the evidence upon which each individual school modifies its instructional program. Some of these districts were also developing and using CRTs, but these tests did not play a prominent part in their instructional renewal program; they were used more as an instructional tool in the classroom rather than as a tool for school-site decision-making.

The school districts using a centralized CRT system focus on a common district instructional continuum, usually in reading, math, and language arts to which all schools are expected to adhere. The impetus for change comes more from the district level than from the local school. The district also encourages the teaching staff to follow a common instructional methodology when implementing the district's curriculum. Student scores on CRTs are used as the main basis upon which instructional effectiveness is gauged. The CRTs are developed so that they relate to the district's adopted instructional program. NRTs are administered and reviewed but they are used mainly to inform the public of the district's program--they do not play a prominent part in the instructional renewal program.

We do not wish to imply that NRTs are not appropriate for decentralized systems or that CRTs are inappropriate for decentralized systems. We are merely reporting that these were the configurations we observed in our small sample of districts. Likely other mixtures of

these elements have been devised.

Although the decentralized and centralized orientations differ in the locus of change and the types of test that are used, the districts' arrangements share important characteristics such as providing support services to the schools: e.g., an extensive and appropriate in-service component, a well-developed data processing capability, a skilled evaluator and measurement staff.

The districts differ in regard to what they considered the effect of their programs. The two centralized, CRT-system districts pointed to what they considered substantial improvements in pupil achievement as a result of their program. The decentralized districts were less sure of the overall effect of their program on student achievement but cited process changes at the school in evidence of effect. This is understandable since the schools themselves differ in what they are trying to accomplish; and these diverse intentions do not lend themselves to more standard yardsticks of progress. Of course, it may be that it takes longer to see the effects of a decentralized program than a more centralized one. We are not yet prepared to offer reasons for, or to assess the differences in the effectiveness of the two approaches. The districts themselves were not presently examining what might be considered unintended or unexpected side effects, e.g., heightened or lowered teacher morale, increased or decreased community support.

Question 3. What are the potential contributions this research has for school improvement?

There is a substantial public and professional "crisis of confidence" in the public schools' ability to adequately educate their pupils--

especially in the basic skills. Increasingly, districts are realizing the limitations of methods of school improvement built on piecemeal approaches, such as untargeted in-service training programs, or new testing programs, or adopting and implementing externally funded projects. These activities, however well-intentioned, simply were not reversing the declining test scores.

Some districts, such as our sample districts, are now seeking more comprehensive and integrated approaches to developing better teaching and learning. One such approach involved connecting the school district's testing and evaluation activities with on-going discussions about how to chart the district pupils' achievement, assess the effects of various instructional strategies, revise those strategies, and use subsequent data to re-assess. We believe this systematic approach will be increasingly tried by other districts. While we think that each district will have to evolve an approach that is appropriate to its particular context and needs, it seems logical that districts beginning to consider this approach can learn a great deal from the experience of these "pioneer" districts. They can learn of the various strategies that have been tried, the specific components (such as CRTs) that have been developed, and the kinds of barriers that have been encountered. Enlightened by the experience of those who have preceded them, these "newer" districts can, perhaps, reduce the time and cost necessary to implement such a system.

Our sample districts have been deeply involved in developing these programs and this has made it difficult for them to step back and take a comprehensive and somewhat detached view of their efforts. What is

more, they do not have the opportunity to compare their efforts with those of other districts that are developing a similar linkage strategy.

We see ourselves as providing two research-related services; first, as observers and recorders of what these districts are doing, so as to subsequently create from their synthesized experiences technical assistance materials for districts wishing to follow this linkage strategy as a means of improving pupil achievement; and second, as analysts of this process, we seek to understand the configuration of human, organizational, political, and technical elements that are associated with the implementation of this linkage strategy so as to contribute to the growing school improvement literature.

With regard to our technical assistance and development role, we realize that the linkage arrangements that our sample districts are developing are unique to each setting and that they cannot be "packaged" and exported to other districts. Nonetheless, there are likely portions of these arrangements that can provide guidance to other districts. The things these sample districts have learned about the process will likely be of considerable interest to those who want to embark on this strategic course. During the last year of this project, we will be working with several districts and helping them begin to design and implement such a program.

With regard to our research/analytical role, we see as a major contribution the bringing together of the research literatures from several fields, e.g., evaluation, testing and curriculum, and organizational theory as a means of gaining insights into the dynamics of this linking

process in school districts. Since these literatures have historically been developed in isolation from each other, our research provides a unique vehicle for gaining a better understanding of their interrelationships. This kind of theory/practice synthesis seems to us to be a necessary step if we are going to be able to fashion research and conceptual work into tools useful for working on the pressing problems facing public education today.

Bibliography

David, J. Local uses of Title I evaluation. Palo Alto, California: Stanford Educational Policy Research Center, 1978.

Lortie, D. C. School teacher. Chicago, Ill.: University of Chicago Press, 1975.

Lyon, C. D., Doscher, L., McGranahan, P., and Williams, R. Evaluation and school districts. Los Angeles: Center for the Study of Evaluation, UCLA Graduate School of Education, 1978.

Weick, K. E. Educational organizations as loosely coupled systems. Administrative Science Quarterly. 1976, 21, 1-19.

Zucker, L. G. Institutional structure and organizational processes: The role of evaluation units in schools. In Evaluation in school districts: Organizational perspectives (CSE Monograph #10). A. Bank & R. C. Williams (eds.). Los Angeles: University of California, 1981.